# Implementation Of Decision Tree Method To Predict Customer Interest In Internet Data Packages

Rezki Surya[1], Muhammad Halmi Dar[2*], Fitri Aini Nasution[3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara Indonesia
[*]Corresponding Author:
Email: mhd.halmidar@gmail.com

***Abstract.***

*The use of smartphones is on the rise, making internet packages vital in everyday life. In Indonesia, mobile operators offer a variety of data packages to meet customer needs. Understanding preferences is important to making the right decisions when providing products that meet customer needs. However, not all packages are suitable for all customers. Telkomsel is one of the providers that can deliver consistent signals and a wide range of data packages, but it still has a relatively high price. The study uses decision tree methodology to analyze citizens' preferences about Telkomsel services, comparing the relative cost of data packages with other services. This research will use surveys with samples from different communities to determine the representativeness of the results and provide strategic recommendations for Telkomsel to improve customer satisfaction. The research methods employed in this study included data collection, preprocessing, data division, model design, prediction results, and result evaluation. The results showed accuracy levels of 98.7%, precision of 100%, recall of 98.7%, specificity of 100%, and an F1-score of 99.3%. This study demonstrates the effectiveness of the decision tree model in predicting customer interest in Telkomsel services. Despite some limitations, the findings provide valuable insights that Telkomsel can use to develop more effective marketing strategies.*

***Keywords:*** *Data Packages, Decision Tree, Internet, Prediction and Telkomsel.*

## I. INTRODUCTION

Internet data packages are services offered by telecommunications service providers to provide Internet access to users via a mobile network [1]. These services allow users to remain connected to the Internet anywhere and at any time without having to rely on a WiFi network. Currently, many data packages with different prices are available to choose from according to user needs and budgets. Telecom providers like Tri, Telkomsel, Axis, Smartfren, IM3, XL, and others offer a wide range of data packages that cover speed, quota, and duration. Users can select the data package that best suits their usage patterns by choosing daily, weekly, or monthly. Prices vary, from the very cheap one for basic use to the more expensive one for greater data speed and volume. In addition, telecommunications service providers often offer promotions and exclusive offers to attract new customers. For example, Telkomsel may offer data packages including additional night quotas, while XL may offer additional quotas for specific applications. For those who need unlimited internet connections, Smartfren offers unlimitated data packages. Users can easily choose the data package that best suits their internet needs, be it for entertainment, education, everyday life, or work. Although there are many data package options available, Telkomsel often charges higher prices compared to other service providers such as Tri, Axis, Smartfren, IM3, and XL. Users with limited budgets or who want to gain more value from data packages may consider these higher prices. Some people who use the Telkomsel card say that the Telkomsel network is still good, even though the data package is more expensive than other service providers.

They believe Telkomsel has more stable connections, higher speeds, and wider network coverage. Users in remote or rural areas often say that Telkomsel is the only provider that can deliver a consistent signal. Furthermore, other service providers have struggled to maintain network stability in the region. The user experience demonstrates that Telkomsel services remain uninterrupted and can manage substantial data volumes, particularly during peak hours. A reliable network quality is crucial for those who rely heavily on an internet connection for work, education, or entertainment purposes, even though they have to pay more. Therefore, although there are cheaper data package options, many users remain loyal to Telkomsel because they believe that better network performance and reliability equal higher costs.With increased smartphone usage, internet data packages are becoming more important in everyday life. With Indonesia's increasing use of the Internet, mobile operators are competing to offer a variety of Internet data packages that are interesting to customers [2]. It is crucial for mobile operators to understand customer preferences and interests in Internet data packets to make informed decisions and provide tailored products to meet customer needs [3]. However, not all such internet data packs match customer needs and preferences. In order to increase customer satisfaction and sales, it is important for mobile operators to be able to predict customer interest in internet data packages. Therefore, analyzing and predicting customer interest patterns in Internet data packets becomes crucial for the mobile operator to develop and offer products that can meet customer needs and preferences more effectively.

Several previous studies have used various methods to predict customer interests. A previous study used the Naive Bayes algorithm to predict customer interest in purchasing internet data packages [4]. The study utilized machine learning techniques to analyze customer demographic data, usage patterns, and historical purchasing behavior to identify key factors that influence customer demand for data packets. The study was able to develop models that can accurately predict customers' interests and preferences, allowing mobile operators to better tailor their product offerings to meet target market needs by applying the NaIVE Bayes classification. Additionally, companies have used clustering techniques like K-Means to group customers based on their purchasing patterns, aiding in their understanding of market segments and customer preferences [5]. By applying these unsupervised learning techniques, researchers were able to identify different customer segments with different consumer preferences and behaviors throughout the year. It allows companies to better understand seasonal fluctuations in customer demand and adjust their product offerings, promotions, and inventory management strategies to more effectively meet the growing needs of their target market segments. The knowledge gained from K-Means cluster analysis allows firms to optimize their sales and marketing efforts, which leads to increased customer satisfaction and, ultimately, increased revenue and profitability [6]. A previous study used data science techniques to analyze insights from supermarket sales data to optimize customer purchasing behavior. By applying sophisticated analysis and machine learning methods to supermarket selling data, researchers could discover valuable patterns and trends that help retailers better understand customer preferences, buying habits, and factors that influence their purchasing decisions.

This allows the supermarket to adjust its product range, prices, promotions, and other marketing strategies more effectively to meet the growing needs and demands of its target customer segment, ultimately leading to improved customer satisfaction, loyalty, and overall sales performance [3]. Naive Bayes's method successfully classified 10 product data with 858 transaction data as training information and another 10 product data with 115 transaction data as test data in a different study [7]. According to the analysis of training data and all test data, the truth ratio is 8–10, which means 8 out of 10 products have true predictive values. If the accuracy value of this prediction is 80%, it is expected to provide a more favorable policy consideration for the sales business. This is achieved by calculating the customer's purchase interest based on the sales data that meets the specified terms and labels. Research results conducted by [8] show that customers have demand from all providers, including Telkomsel, Indosat, and Axis. One of the results of the analysis carried out using the C.45 algorithm with the Rapidminer tool is that the result of the decision tree shows consumer interest through price. Conversely, the C.45 algorithm, also known as the decision tree, yields an accuracy rate of 94.67%. However, previous studies have not specifically tested the use of the decision tree method to predict customer interest in internet data packages. As a result, the study will focus on how to apply the decision tree method in that context.An increasingly popular data mining technique, the decision tree method, creates predictive models based on decision rules [9].

This powerful machine learning algorithm works by separating input data into several sub-sets, each distinguished by a specific attribute value. This process forms a hierarchical structure like a tree consisting of nodes and branches that visually represent a variety of potential decisions and outcomes [10]. Each internal node in this tree corresponds to an attribute, with the branches derived from that node representing a different rule of decision based on that attribute's value. The leaf node at the end of these branches then marks the predicted classes or final target results [11]. The main advantage that makes decision trees so convincing is their inherent ability to produce models that are highly understandable and interpretable, allowing users to easily identify the most influential factors that drive the predicted results [12], [13]. This transparency and ease of interpretation are crucial, as they allow domain experts and non-technical stakeholders to better understand the logic underlying model predictions. Many applications, such as classification, regression, and business decision-making, highly value the accuracy, flexibility, and interpretability of decision tree methods [14]. In a recent study, researchers used the Decision Tree algorithm to predict customer preferences for a variety of product offers [15].

By applying this technique to customer sales data, the researchers were able to uncover hidden patterns and relationships that enabled companies to better tailor their product portfolios, promotions, and supply management strategies to the growing needs and desires of the target market segment [16]. This, in turn, led to increased customer satisfaction, increased sales, and improved overall business performance. The aim of this study is to apply the decision tree method to determine how much public interest there is in Telkomsel services, even though the price of the data package is relatively high compared to other service providers. By analyzing variables such as price, network availability, and network quality, we expect the research to provide a clearer picture of users' preferences in choosing telecommunications services. We will conduct surveys involving samples from different communities to ensure the results are representative. This analysis not only shows the level of public interest in Telkomsel data packages but also helps determine the most potential market segments. Therefore, we expect this research to provide strategic recommendations that benefit Telkomsel by improving their services to meet customer needs and expectations, thereby increasing customer attractiveness and satisfaction.

## II. METHODS

Due to its ability to identify and visualize patterns in complex datasets, data mining uses the decision tree method to determine public interest in Telkomsel data packages. The decision tree will help in this research by breaking the data sets into several subsets based on important features such as name, card type, price, network availability, and network quality preferences. The tree structure will show how different segments of the community make decisions when choosing Telkomsel data packages. Therefore, researchers can identify which components are most influential in attracting public attention to this service. The first step is data selection.

This procedure collects the data needed for the research. The procedure will first collect the data for use, but it cannot be used directly. We must select and validate the data before use. The first step in this process will be to preprocess the data for use in this research. We will use two sets of data: training data and test data. We use the exercise data to forecast public interest in Telkomsel's data package. We use the sample data, known as the test data, to make predictions. The model design phase, the second phase, involves constructing a predictive model to forecast the relevant data in the telecommunications data bundle. The prediction results, obtained from a prediction made using a model already designed using the decision tree method, represent the third phase of the process. The final phase will be the result evaluation, which will display the confusion matrix by showing the accuracy, precision, recall, specificity, and F1-score values.

## III. RESULT AND DISCUSSION

We present the research results systematically in accordance with the previously described research phases, using the decision tree method to predict customer interest in internet data packages.

**Table 1.** Dataset

| Respondents | Providers | Price | Network Availability | Network Quality | Category |
|---|---|---|---|---|---|
| Respondent 1 | Telkomsel | Expensive | Not Available | Not Good | Not Interest |
| Respondent 2 | Telkomsel | Cheap | Not Available | Not Good | Not Interest |
| Respondent 3 | Telkomsel | Cheap | Available | Good | Interest |
| Respondent 4 | Tri | Expensive | Not Available | Not Good | Not Interest |
| Respondent 5 | Smartfren | Expensive | Not Available | Not Good | Not Interest |
| Respondent 6 | Telkomsel | Cheap | Available | Good | Interest |
| Respondent 7 | Tri | Cheap | Available | Good | Not Interest |

| | | | | | |
|---|---|---|---|---|---|
| Respondent 8 | Telkomsel | Cheap | Available | Good | Interest |
| Respondent 9 | Telkomsel | Cheap | Available | Good | Interest |
| Respondent 10 | IM3 | Expensive | Not Available | Not Good | Not Interest |

Table 1 contains sample data for predicting customer interest in internet data packages. The data set comprises 100 rows, each with 6 columns. We use a total of 19 data points for training, and 81 data points for testing. The Respondents column enumerates each respondent's identification, assigning each respondent a unique number. Providers is a column that lists the Internet service providers used by each respondent. The aforementioned service providers include Telkomsel, Tri, Smartfren, IM3, Axis, and XL. Price is a column that lists the price categories of data packages used by respondents, namely "expensive" or "cheap". The column "Network Availability" lists network availability in the respondent area, with a value of "Available" or "Not Available". A category is a target column or label that indicates a customer's interest in an Internet data package, with a value of "Interest" or "Not Interest".

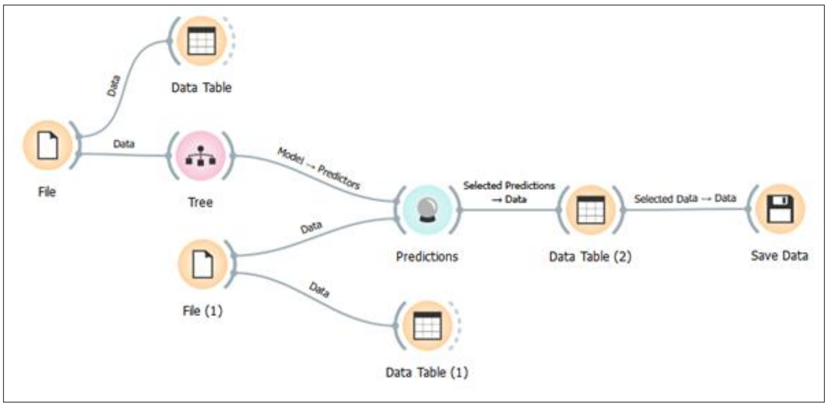


**Fig 1.** Design Model

Figure 1 is a model design flow diagram of the decision tree method for predicting customer interest in internet request packages. This diagram describes steps in the process of making predictive models using the decision-tree method. File: This step is the process of retrieving data from the source file. We will use this data to build a predictive model. Data Table: The data table receives data from the first file. The subsequent process uses this table as a data structure. Tree: This step uses the decision tree algorithm to build a predictive model. We train this model using the data from the previous table. The decision tree model will learn from the data to determine the rules for predicting customer interests. Predictions: We then use the trained decision tree model to make predictions based on new data. This node takes data from the table and generates the corresponding prediction. Data Table (2): The model decision tree inserts the generated predictions into the new data table. This table contains the predictive data that the model generates. Save Data: The final step is to store the prediction data in a file. You can use this stored data for further analysis or to report results. File (1) is an additional data source that may contain new data or validation data. The prediction process also uses this data. Table Data (1) We insert the data from this additional file into another data table and use it alongside the data from the primary file.

| | | Predicted | | |
|---|---|---|---|---|
| | | Interest | Not Interested | ∑ |
| Actual | Interest | 75 | 0 | 42 |
| | Not Interested | 1 | 5 | 50 |
| | ∑ | 42 | 50 | 92 |

**Fig 2.** Confusion Matrix

Based on a total of 81 test data points, Figure 2 displays a confusion matrix to evaluate the performance of a classification model predicting customer interest in internet data packages. Interest is defined as the number of customers who are actually interested in an Internet data package. Not Interested refers to the total number of clients who are not actually involved in an internet data packet. The confusion matrix shows true positive (TP) results: as many as 75 predicted customers are interested and are actually interested. False Positive (FP): as much as 0 predicated clients are interested but are actually not interested. False Negative (FN): as many as 1 predicted customer is not interested but is actually interested. True negative (TN): As many as 5 customers are not interested and are actually not interested. Number of customers that are interested: 75 + 1 = 76 customers. Number of clients that are not interested: 0 + 5 = 5 customers.

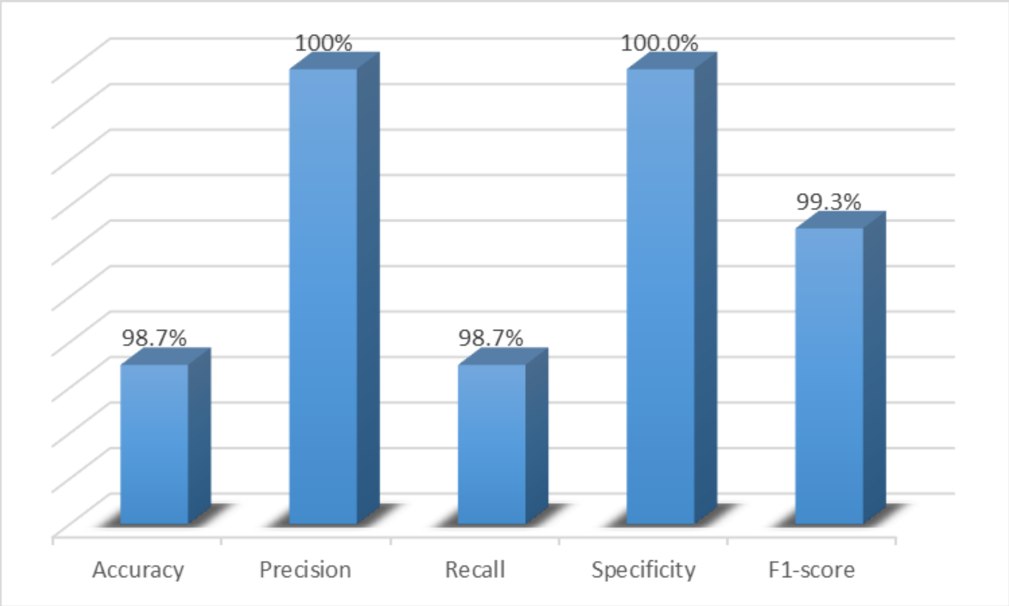


**Fig 3.** Evaluation Matrix

Figure 3 shows the results of an evaluation of the Decision Tree model's performance in predicting customer interest in internet data packages. Accuracy (98.7%) measures the proportion of correct predictions (both positive and negative) to the total prediction. The result indicates that 98.7 percent of the model prediction corresponds to actual data. Precision100% measures the true positive prediction ratio of all positive predictions. According to actual data, a precision value of 100% indicates that the model actually attracts all interested customers. Recall: The proportion of actual positive cases correctly predicted by the model was measured at 98.7%. A recall value of 98.7% indicated that the model correctly identified 98.7% of the actual customers who were interested. Specificity: 100% measures the proportion of actual negative cases correctly predicted by the model.

A 100% specificity value indicates that all customer predictions are not really based on actual data. The F1 Score of 99.3% is the harmonic mean of precision and recall, providing a balanced picture of both metrics. The F1-Score value of 99.3% demonstrates the excellent performance of the model, with a good balance between precision and recall.The decision tree has advantages in terms of readability and interpretation. The decision tree structure allows researchers and stakeholders to understand the logic behind the predictions made. In a business context, the urgent need for transparency and decision justification makes it essential. This method is capable of dealing with data complexity involving a variety of variables that affect customer interests. It is important to keep in mind that customer decisions to choose a data package depend not only on one factor but also on a combination of several variables, such as price, quality of service, promotion, and brand perception. The decision tree model, with an accuracy of 98.7%, demonstrates that nearly all predictions align with actual data. It shows that the model is very accurate in predicting customer interests.

## IV. CONCLUSION

With a 98.7% accuracy rate, the decision tree model is very accurate in predicting customer interest in Telkomsel data packages. A precision value of 100% indicates that all predictions about customers attracted by this model are actually true. There were no false positive predictions (false positives). The model successfully predicts almost all actual customers interested, as evidenced by recalls of 98.7%. There are only a few errors in the form of false negative predictions. A 100% specificity value shows that all non-interested customer predictions are completely uninterested. No fake negative forecasts. (False Negatives). The F1 score of 99.3% shows an excellent balance between precision and recall, confirming the model's effective and efficient performance in predicting customer interests.This study's decision tree model accurately predicts customer interests, indicating that the price factor does not entirely stifle public interest in Telkomsel services. The model decision tree has proven to be effective in analyzing data and predicting customer interest, providing accurate and reliable results. These results demonstrate the potential of the decision tree model in various business applications, particularly in the domains of consumer behavior prediction and market analysis. Further research may consider additional variables that may affect customer interests, such as quality of service, promotion, and customer service. You can get a better idea of how well the decision tree model works by comparing it to other classification methods, such as random forest, support vector machine (SVM), or neural networks.

**REFERENCES**


[1] A. Lia Hananto, M. Guntur, and B. Priyatna, "Perencanaan Arsitektur Enterprise pada Layanan Telekomunikasi Digital Telkomsel By.U menggunakan Zachman Framework," ***J. Account. Inf. Syst. (AIMS***, vol. 7, no. 1, pp. 9–16, 2024, doi: https://doi.org/10.32627/aims.v7i1.916.

[2] K. AP, "Analysis Implementation Analytical Hierarchy Process Method and Weighted Product for Ranking Internet Package Selection System," in Procedia of Engineering and Life Science, Universitas Muhammadiyah Sidoarjo, 2021. doi: doi.org/10.21070/pels.v2i0.1173.

[3] M. Harahap, F. Rozi, Y. Yennimar, and S. D. Siregar, "Analisis Wawasan Penjualan Supermarket dengan Data Science," Data Sci. Indones., vol. 1, no. 1, pp. 1–7, 2021, doi: 10.47709/dsi.v1i1.1173.

[4] D. R. Prehanto, A. D. Indriyanti, K. D. Nuryana, S. Soeryanto, and A. S. Mubarok, "Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token," ***J. Phys. Conf. Ser***., vol. 1402, no. 6, p. 66069, 2019, doi: 10.1088/1742-6596/1402/6/066069.

[5] Y. D. Putra, M. Sudarma, and I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," ***J. Hari. Reg***., vol. 20, no. 2, 2021, doi: https://doi.org/10.24843/MITE.2021.v20i02.P03.

[6] T. N. Suharsono and M. Akbar, "Prediksi Minat Konsumen Sesuai Musim Menggunakan Algoritma K-Means Pada Perangkat Lunak Pemesanan Tempat Dan Makanan Online," ***Infotronik J. Teknol. Inf. dan Elektron***., vol. 2, no. 2, p. 124, 2017, doi: 10.32897/infotronik.2017.2.2.39.

[7] A. S. Mubarok, D. R. Prehanto, and M. Ali, "Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma Naive Bayes Classifier," ***Inov. J. Ilm. Inov. Teknol. Inf***., vol. 3, no. 2, 2019, [Online]. Available: https://ejournal.unhasy.ac.id/index.php/inovate/article/view/737

[8] H. Susana, S. Khofidoh, and A. Al Afghani, "Implementasi Data Mining Metode C.45 untuk Memprediksi Peminat Kuota Internet pada Masa Covid 19," MEANS (Media Inf. Anal. dan Sist., vol. 6, no. 2, pp. 155–157, 2022, doi: 10.54367/means.v6i2.1513.

[9] N. Pandiangan, M. L. C. Buono, and S. H. D. Loppies, "Implementation of Decision Tree and Naïve Bayes Classification Method for Predicting Study Period," ***J. Phys. Conf. Ser***., vol. 1569, no. 2, p. 22022, 2020

[10] M. A. Hafeez, M. Rashid, H. Tariq, Z. U. Abideen, S. S. Alotaibi, and M. H. Sinky, "Performance Improvement of Decision Tree: A Robust Classifier Using Tabu Search Algorithm," Applied Sciences, vol. 11, no. 15. 2021.

[11] M. Méndez, M. G. Merayo, and M. Núñez, "Machine learning algorithms to forecast air quality: a survey," *Artif. Intell. Rev*., vol. 56, no. 9, pp. 10031–10066, 2023, doi: 10.1007/s10462-023-10424-4.

[12] Z. Zhang, "Applications of the Decision Tree in Business Field BT - Proceedings of the 2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021)," Atlantis Press, 2021, pp. 926–929. doi: 10.2991/assehr.k.211209.151.

[13] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," ***J. Appl. Sci. Technol. Trends***, vol. 2, no. 01 SE-, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[14] L. Cañete-Sifuentes, R. Monroy, and M. A. Medina-Pérez, "A Review and Experimental Comparison of Multivariate Decision Trees," IEEE Access, vol. 9, pp. 110451–110479, 2021.

[15] D. Ignatov and A. Ignatov, "Decision Stream: Cultivating Deep Decision Trees," in *2017* IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017, pp. 905–912.

[16] M. J. Awan, M. S. M. Rahim, H. Nobanee, A. Yasin, O. I. Khalaf, and U. Ishfaq, "A Big Data Approach to Black Friday Sales," Intelligent Automation \& Soft Computing , vol. 27, no. 3. 2021.